Inv Ed Med. 2016;xxx(xx):xxx–xxx

Investigación en Educación Médica
http://riem.facmed.unam.mx

ARTÍCULO ORIGINAL

Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad

Olivia Espinosa-Vázquez[a,*], **Adrián Martínez-González**[b], **Melchor Sánchez-Mendiola**[b] e **Iwin Leenen**[c]

[a] *Facultad de Odontología UNAM, Ciudad de México, México*
[b] *Coordinación de Desarrollo Educativo e Innovación Curricular UNAM, Ciudad de México, México*
[c] *Instituto Nacional para la Evaluación de la Educación, Ciudad de México, México*

Recibido el 18 de abril de 2016; aceptado el 6 de septiembre de 2016

PALABRAS CLAVE
México;
Competencia clínica;
Teoría de la generalizabilidad;
Examen clínico objetivo estructurado (ECOE);
Odontología

Resumen

Introducción: Diversos estudios han analizado al examen clínico objetivo estructurado (ECOE) en Odontología para acumular evidencia de validez al utilizarlo como una herramienta de evaluación de la competencia clínica en los estudiantes. En este estudio introdujimos un ECOE diseñado en Odontología y se discuten los resultados del análisis desde la perspectiva de la teoría de la generalizabilidad, utilizando datos obtenidos de una aplicación del examen.

Método: Se realizó un estudio observacional y transversal en la Facultad de Odontología de la UNAM. Participaron 120 estudiantes en un ECOE diseñado *ex profeso* en un circuito de 18 estaciones con duración de 6 min cada una, en el contexto de un curso de Odontopediatría del cuarto año de la licenciatura en Cirujano Dentista. Un análisis basado en la teoría de la generalizabilidad, con evaluadores y estaciones considerados como facetas, identificó las principales fuentes de variabilidad en los datos.

Resultados: La media (y desviación estándar) global de las calificaciones en el examen corresponde a 44% (7%), con las medias por estación variando entre el 23 y el 66%. El estudio de generalizabilidad mostró que la faceta correspondiente a los evaluadores explicó una parte significativa (13%) de la variación en los resultados por estación, más que la competencia clínica de los sustentantes (6%). En el estudio de decisión se encontró un coeficiente de generalizabilidad relativo de 0.63 y absoluto de 0.55.

Conclusiones: A la luz de los coeficientes de generalizabilidad relativamente bajos en el estudio de decisión, es importante analizar más allá el desarrollo del ECOE-O para minimizar el efecto de las fuentes que introducen varianza irrelevante al constructo en los resultados;

* Autor para correspondencia. Universidad 3000, Coyoacan, Copilco Universidad, 04360 Ciudad de México, D.F., México.
Correo electrónico: oliviaedunam@live.com.mx (O. Espinosa-Vázquez).
La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

http://dx.doi.org/10.1016/j.riem.2016.09.001

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

especialmente, se requiere revisar y ajustar las estaciones, así como calibrar mejor a los profesores para homogeneizar los criterios de evaluación.

KEYWORDS
Mexico;
Clinical competence;
Generalisability theory;
Objective structured clinical examinations (OSCE);
Dentistry

Analysis of an objective structured clinical examination in dentistry using generalisability theory

Abstract
Introduction: Various studies have examined Objective Structured Clinical Examinations (OSCEs) in Dentistry in order to accumulate validity evidence for their use as an assessment tool of clinical competence in students. In this article, a newly designed OSCE in Dentistry (OSCE-D) is introduced and discussion is presented on the results of an analysis from the perspective of generalisability theory using data obtained from an application of the examination.
Method: An observational and cross-sectional study was conducted in the Faculty of Dentistry at UNAM. One hundred and twenty pre-graduate students participated in an OSCE that consisted of 18 stations, with a duration of 6 min each, in the context of a fourth-grade Paediatric Dentistry course. An analysis based on generalisability theory, with raters and stations being considered as facets, identified the main sources of variability in the data.
Results: The overall mean (and standard deviation) of the OSCE score, across participants and stations, was 44% (7%), with the station means varying between 23% and 63%. The generalisability study showed that the facet of the raters explained a significant portion (13%) of the variance in the station results, which was more than the clinical competence of the participants (6%). The decision study produced a generalisability index of 0.63 and a dependability index of 0.55.
Conclusions: In view of the rather low reliability indices from the decision study, it is important to make a further analysis of the OSCE-D so as to minimise the effect of sources that introduce construct-irrelevant variance into the results. In particular, an adjustment of the stations may be required, as well as a better standardising in the use of evaluation criteria by the raters.

Introducción

La investigación de la evaluación de la competencia clínica en el área de la salud ha brindado aportaciones para identificar fortalezas y debilidades en el diseño de los instrumentos utilizados en aspectos diagnósticos, formativos y/o sumativos[1-6].

Para valorar la competencia clínica, se utiliza una combinación de herramientas integrales tales como la evaluación de 360°, el examen ante paciente real[7,8], el informe de prácticas y el examen clínico objetivo estructurado (ECOE). Este último fue introducido por primera vez en el campo de la educación médica en 1975 por Harden et al.[9]; consiste en que los examinados rotan alrededor de un circuito de estaciones en las que deben desarrollar una tarea clínica relacionada con la disciplina o área por evaluar, con un tiempo determinado en cada una. Cada estación debe establecer un objetivo, un lugar donde desarrollarse, el material para el estudiante, una hoja para el evaluador y otra para la captura de la información[10-12]. Las estaciones del ECOE pueden clasificarse en 2 tipos: 1) dinámicas, donde el estudiante tiene una tarea clínica por realizar con un paciente estandarizado o en un simulador, y 2) estáticas, donde tiene que responder cuestionamientos sobre la base de la información que ha obtenido en la misma estación[13].

En México, el ECOE ha sido empleado desde 1997 en la Facultad de Medicina de la UNAM como herramienta de evaluación formativa y sumativa[14], y no se ha reportado su uso en el ámbito odontológico, aun cuando se introdujo en esta disciplina a nivel internacional en ese mismo año[1,15]. El ECOE en odontología ha sido utilizado recurrentemente para evaluar habilidades clínicas con simuladores[16,17]. Entre sus ventajas se encuentran que los examinadores pueden decidir con antelación qué es lo que va a ser evaluado y posteriormente diseñar el examen bajo un objetivo determinado; controlar el contenido y las complejidades de la prueba al explorar un amplio rango de habilidades; asimismo puede utilizarse con un gran número de estudiantes y brindarles realimentación[18,19].

El ECOE, como cualquier instrumento de evaluación de habilidades clínicas, debe diseñarse con rigor metodológico para la correcta interpretación de sus resultados. Entre las cualidades que deben sustentar su uso se encuentra la validez, referida a la evidencia presentada para apoyar o refutar las inferencias a partir de los resultados de determinada evaluación[20]. Los modelos actuales consideran a la validez como un concepto unitario que contempla múltiples recursos de evidencia clasificados en 5 grandes fuentes: contenido, procesos de respuesta, estructura interna, relación con otras variables y consecuencias[20-22]. Para este reporte,

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

se describe el diseño de un ECOE en odontología y, a partir de los resultados, se analiza su confiabilidad, que es una de las evidencias de validez relacionada con la estructura interna del instrumento.

Desde la teoría clásica de los tests (TCT), la definición de la confiabilidad parte de la idea de que la puntuación observada en una prueba es el resultado de la puntuación verdadera de la persona y un error que afecta la medición de una forma no sistemática[23]; desde esta perspectiva, se cuantifican 2 componentes de varianza en las puntuaciones observadas: la varianza verdadera y la varianza de error. La confiabilidad indica la proporción de la varianza verdadera en la varianza total de las puntuaciones observada; varía de 0 a 1, y los valores más cercanos a 1 indican una confiabilidad más elevada.

Se han propuesto diversas formas para estimar la confiabilidad desde la TCT: a través de la correlación test-retest, la correlación entre formas paralelas y el índice más reportado y utilizado, el alfa de Cronbach[24]; se calcula a partir de una prueba administrada en una sola ocasión y cuantifica la consistencia interna entre los componentes del examen (por ejemplo, las estaciones de un ECOE). Brown et al.[25] hallaron un alfa de Cronbach de 0.68 en un ECOE en odontología en estudiantes del cuarto año, diseñado para evaluar la competencia clínica de los estudiantes y proporcionarles realimentación; Gerrow et al.[26] analizaron la confiabilidad del ECOE, que es parte del examen de certificación de dentistas en Canadá, y encontraron coeficientes alfa de 0.69 a 0.74; Näpänkangas et al.[27] realizaron un estudio para evaluar la correlación entre los resultados de un ECOE en odontología y la evaluación clínica de los estudiantes a lo largo de su formación profesional, así como para probar la confiabilidad de las estaciones; obtuvieron un alfa de Cronbach de 0.94.

En México no se encontraron reportes con respecto al ECOE en el ámbito odontológico, pero se reportó la confiabilidad de un ECOE en estudiantes de medicina con el alfa de Cronbach en un estudio pre-postest[14]. Los índices reportados fueron 0.62 y 0.64, respectivamente.

Cronbach y colegas desarrollaron la teoría de la generalizabilidad como una extensión de la TCT, que se diferencia de esta en el sentido de que distingue entre múltiples fuentes de error. En particular, dicha teoría consiste en 2 estudios: a) el estudio de generalizabilidad (estudio G), el cual estima la contribución de (las habilidades evaluadas de) los sustentantes, junto con la de las fuentes de error consideradas (facetas), a la varianza en las puntuaciones observadas, y b) el estudio de decisión (estudio D), que utiliza la información proporcionada por el estudio G para evaluar el diseño de la medición para un propósito particular y que permite estimar los índices de generalizabilidad relativos y absolutos asociados (llamados *generalizability index* y *dependability index* en la literatura anglosajona). Ambos índices toman valores entre 0 y 1 y funcionan similar a un coeficiente de confiabilidad; el primero sirve cuando el objetivo es realizar decisiones relativas, comparando los estudiantes entre sí con base en su desempeño, y el segundo cuando se toman decisiones referidas a un criterio, donde se contempla el nivel del desempeño de un sustentante independientemente del desempeño de los otros (comparándolo con algún estándar externo, por ejemplo)[24,28,29].

De entre los estudios analizados a la luz de esta teoría destaca el de Schoonheim-Klein et al.[30], quienes estudiaron la confiabilidad de un ECOE en odontología administrado en múltiples días, y evaluaron el número de estaciones necesarias para una decisión confiable tanto para decisiones relativas como absolutas; para decisiones relativas reportaron un coeficiente de generalizabilidad de .62, y para las absolutas, de .54. Con base en sus resultados, el mínimo de estaciones necesarias para decisiones relativas es de 12, y para absolutas es de 17. Asimismo, Eberhard et al.[31] evaluaron la confiabilidad, la validez, la factibilidad y el efecto del número de estaciones de un ECOE en odontología en un curso propedéutico; los índices de generalizabilidad relativo y absoluto obtenidos fueron 0.75 y 0.69, respectivamente, y 14 son las estaciones mínimas requeridas para garantizar una confiabilidad mínima de 0.8, según este estudio.

Bergus y Kreiter[32] buscaron determinar la generalizabilidad de la síntesis de los puntajes acumulados de un ECOE formativo con casos distribuidos a través de 5 prácticas clínicas principales, durante el tercer año del pregrado en medicina de 2 generaciones. Para el estudio G, estos autores reportaron el 9.7% del puntaje de la varianza originada de los estudiantes, el 3.1% de la interacción práctica clínica-estudiante y el 87.2% de la interacción anidada estudiante-caso dentro del efecto práctica clínica. A partir de los resultados del estudio G, el estudio D reportó un índice relativo de 0.63 para los estudiantes que completaron 3 casos en cada una de las 5 diferentes prácticas clínicas.

En México, Trejo et al.[33] reportaron un índice de confiabilidad a través de la teoría de la generalizabilidad que varió entre 0.81 y 0.93 en 7 generaciones en las que se aplicó el ECOE en medicina.

En la presente investigación se describe el diseño de un ECOE en odontología (ECOE-O) en un curso de Odontopediatría en la etapa preclínica, así como el análisis de la confiabilidad de los resultados con la teoría de la generalizabilidad, con la finalidad de identificar diversas fuentes que influyen en la evaluación de la competencia clínica de los estudiantes en el ECOE-O, y con ello optimizar el diseño del instrumento.

Método

Participantes

Se realizó un estudio observacional y transversal con 120 estudiantes (90 mujeres y 30 hombres) de la Facultad de Odontología de la Universidad Nacional Autónoma de México que cursaban la etapa preclínica de la asignatura de Odontopediatría del cuarto (penúltimo) año de la licenciatura de Cirujano Dentista, y que estaban inscritos en uno de los 4 grupos de esta asignatura que se seleccionaron para participar en este estudio, excluyendo a los recursadores.

Instrumento

El ECOE-O consistió en un circuito de 18 estaciones con duración de 6 min cada una, formato seleccionado al considerar la experiencia de la entidad académica que asesoró el diseño de este examen en el contexto de la formación profesional en el área de la salud en México[4]. Se incluyeron

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

Tabla 1 Estaciones que conformaron el ECOE-O en la asignatura de Odontopediatría en el cuarto año de la licenciatura de Cirujano Dentista de la Facultad de Odontología de la UNAM

Núm.	Tipo	Atributo	Tema	Núm. de ítems
1	D	Habilidades técnicas	Aislamiento absoluto	15
2	D	Habilidades técnicas	Anestesia	20
3	D	Habilidades técnicas	Operatoria dental	13
4	D	Habilidades técnicas	Selladores de fosetas y fisuras	10
5	E	Interpretación y diagnóstico radiográfico	Patología oral	5
6	E	Exploración física y diagnóstico	Desarrollo de la oclusión	13
7	E	Interpretación y diagnóstico radiográfico	Cronología y secuencia de erupción	7
8	E	Exploración física y diagnóstico	Caries por alimentación infantil	5
9	E	Plan de tratamiento	Prescripción farmacológica	7
10	D (PE)	Prevención y promoción de la salud	Métodos de prevención para caries y enfermedad periodontal	16
11	D (PE)	Comunicación y profesionalismo	Abordaje de la conducta	9
12	D (PE)	Exploración física y diagnóstico	Patología oral	18
13	E	Plan de tratamiento	Anestesia	4
14	E	Interpretación y diagnóstico radiográfico	Caries tercer grado	11
15	D (PE)	Prevención y promoción de la salud	Métodos de prevención para caries y enfermedad periodontal	4
16	E	Comunicación y profesionalismo	Abordaje de la conducta	6
17	E	Interrogatorio	Lesiones traumáticas	8
18	D (PE)	Plan de tratamiento	Lesiones traumáticas	6

D: dinámica; E: estática; ECOE-O: examen clínico objetivo estructurado en odontopediatría; PE: paciente estandarizado.

9 estaciones estáticas y 9 dinámicas; de estas, 4 se desarrollaron con simuladores y 5 con pacientes estandarizados (niños y adultos quienes tenían experiencias previas dentales y el ambiente les era familiar). Es importante destacar que 2 de las estaciones seleccionadas (10 y 15) evaluaron el mismo dominio y el mismo tema. Para más detalles, véase la tabla 1.

Para elaborar las estaciones se construyó una matriz de competencias que describía los 7 dominios de la competencia clínica que serían evaluados en los estudiantes, así como los temas que abarca el programa de la asignatura. Los 7 dominios fueron: a) comunicación y profesionalismo; b) interrogatorio; c) exploración física y diagnóstico; d) diagnóstico e interpretación radiográfica; e) plan de tratamiento; f) prevención y promoción de la salud bucodental, y g) habilidades técnicas. Se eligieron con base en una búsqueda de la literatura[27,30,34-37], así como en la opinión de profesores especialistas en Odontopediatría, quienes además fueron capacitados en un taller para diseñar las estaciones y desempeñar su papel como evaluadores en el examen. Antes de su inclusión en el examen final, las estaciones fueron evaluadas en una prueba piloto que se realizó con 40 estudiantes y 22 examinadores en la que se evaluaron aspectos tales como: relevancia del contenido, rol del examinador, realidad de los escenarios, claridad de las instrucciones y de los enunciados en las rúbricas, calidad de los materiales empleados y desempeño de los pacientes estandarizados; la información se recopiló a través de la técnica de grupos focales con estudiantes, de formatos para obtener la información por parte de los examinadores, y de un análisis estadístico simple; con ello se modificaron algunos ítems, se corrigieron instrucciones de algunas estaciones y se optimizó el material utilizado (principalmente radiografías y fotografías impresas).

Procedimiento y evaluación por jueces

Los estudiantes se presentaron al examen distribuidos en 6 turnos durante 2 días. La dinámica del examen se les explicó días previos al mismo y minutos antes de presentarlo. La duración del examen fue de 120 min. Se contó con la participación de 45 evaluadores, quienes fueron asignados a las estaciones con base en su desempeño en el taller de capacitación y en la experiencia que tenían en los temas de la asignatura. El evaluador emitió un juicio sobre el desempeño del estudiante en la estación a través de una rúbrica con una serie de ítems (entre 4 y 20, dependiendo de la estación) que explicitaban criterios objetivo de los dominios de la competencia clínica; para cada uno se eligió entre 4 niveles de desempeño (deficiente, regular, bueno, excelente). En el caso de las estaciones dinámicas con paciente estandarizado, se realizaron 2 juicios adicionales: a) una valoración global (en una escala de 1 a 9) de las habilidades

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

de comunicación interpersonal (HCI), donde 1 representó un desempeño insatisfactorio respecto a la atención personal que brindaba al paciente (saludo, respeto, atención, lenguaje adecuado) y 9 caracterizó un desempeño impecable en los aspectos señalados; y b) un juicio sobre el trato recibido del estudiante al paciente estandarizado evaluado con una escala de 5 a 10, donde 5 era deficiente y 10 excelente.

Análisis

Puntajes por estación

En cada estación dinámica con paciente estandarizado se obtuvo una calificación que contemplaba el 85% del resultado de los ítems de la rúbrica, el 10% de la calificación de la escala HCI y el 5% de la calificación otorgada por el paciente estandarizado, y cuando fueron estaciones dinámicas con simuladores, el 100% de la calificación fue para el resultado de la rúbrica; lo mismo sucedió con las estaciones estáticas. Estos porcentajes fueron determinados por los profesores especialistas ya mencionados, al considerar que los ítems de la rúbrica tenían una significatividad mayor en cuanto a su contenido y estructura para evaluar los dominios.

Análisis desde el marco de la teoría de la generalizabilidad

Una vez obtenida la calificación por estación, se realizó un análisis a través de la teoría de la generalizabilidad. Para el estudio G se consideraron 2 facetas: las estaciones y los evaluadores. Siguiendo la notación en la literatura sobre el tema[38], el diseño para este estudio se denota $p \times (o:t)$, debido a que la faceta de los evaluadores (*observers* en inglés) está anidada en las estaciones (*tasks* en inglés); es decir, cada evaluador está asociado únicamente con una estación. Esto implica que se pueden estimar los componentes de varianza en la siguiente ecuación:

$$\sigma^2_X = \sigma^2_p + \sigma^2_t + \sigma^2_{o,ot} + \sigma^2_{\text{residual}} \quad (1)$$

donde σ^2_X es la varianza total de los puntajes observados y σ^2_p, σ^2_t y $\sigma^2_{o,ot}$ se refieren a la varianza atribuible a los estudiantes, las estaciones y los evaluadores anidados en las estaciones (por lo cual incluye la varianza debida a la interacción entre estaciones y evaluadores), respectivamente. El último componente, $\sigma^2_{\text{residual}}$, reúne la varianza de las fuentes no incluidas en los componentes anteriores (como las interacciones entre estudiantes y estaciones y entre estudiantes y evaluadores, la interacción triple y el error). Al dividir las respectivas varianzas al lado derecho de la igualdad entre la varianza total, se obtiene la contribución proporcional de las distintas fuentes de variación en los puntajes observados.

A partir de los resultados del estudio G, se realizó el estudio D y se obtuvieron los índices relativos y absolutos; asimismo se reporta cómo cambiarían estos índices si se decidiera aumentar o disminuir el número de estaciones y/o el de evaluadores por estación en este estudio.

Consideraciones éticas

La participación de los profesores, estudiantes y pacientes estandarizados fue voluntaria y el estudio no consideraba ningún riesgo para los participantes. En el caso de los estudiantes, se solicitó su consentimiento informado por escrito; asimismo se les informó que el resultado obtenido en estas pruebas no repercutiría en su calificación final del curso, y les sería reportado de manera individual y confidencial.

Resultados

Resultado de las calificaciones globales

La media global de las calificaciones de los estudiantes corresponde al 44% y una varianza del 7%, con las medias por estación que varían entre el 23 y el 66% (fig. 1).

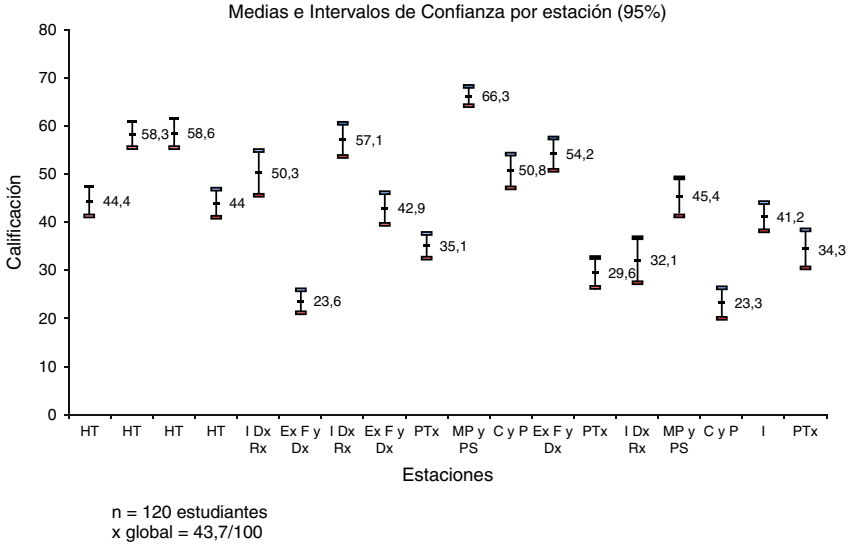

Figura 1 Resultados obtenidos por los estudiantes de forma global y por estación en el ECOE-O. Las barras representan intervalos de confianza. El eje de las abscisas indica las estaciones y el de ordenadas, la media aritmética.
C y P: comunicación y profesionalismo; ExF y D: exploración física y diagnóstico; HT: habilidades técnicas; I Dx RX: interpretación y diagnóstico radiográfico; I: interrogatorio; MP y PS: métodos de prevención y promoción de la salud; PTx: plan de tratamiento.

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

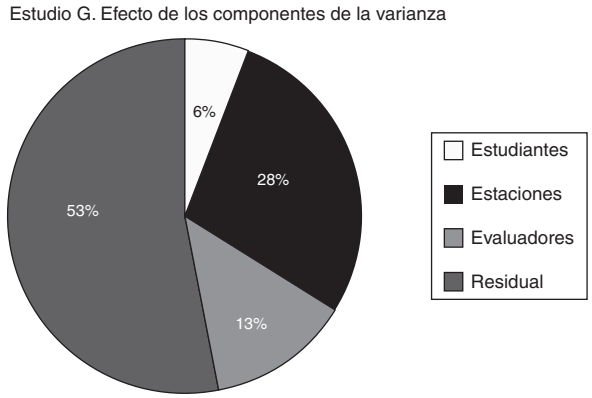

Figura 2 Gráfica del estudio G en la que se muestra el efecto que cada faceta del ECOE-O tiene en la evaluación de la competencia clínica de los estudiantes en la asignatura de Odontopediatría.

Resultados del estudio G

La contribución de las diversas fuentes consideradas en este estudio a la varianza en los puntajes del ECOE (véase la ecuación 1) fue el siguiente: el 6% corresponde al efecto de los estudiantes, el 28% a las estaciones, el 13% a los evaluadores y el 53% corresponde al denominado efecto o error residual (fig. 2). Destaca que la varianza debida a los estudiantes es baja (6%), lo cual implica que, al considerar los puntajes por estación (contrario al puntaje en el examen global, véase el estudio D), las diferencias entre estudiantes respecto de su competencia clínica explican una parte relativamente pequeña de la variación en dichos puntajes. Diferencias en el grado de dificultad entre las estaciones y en el grado de severidad entre los evaluadores influyen más en la variación de los puntajes en las estaciones. Sin embargo, el componente de varianza más importante es el residual, lo cual implica que interacciones entre las fuentes incluidas en el estudio y/u otras fuentes no consideradas tienen un efecto significativo.

Resultados del estudio D

Para el estudio D se contemplaron en un primer paso los índices de generalizabilidad relativo y absoluto para el mismo diseño que el utilizado para el estudio G (18 estaciones con un evaluador anidado en cada estación). El índice relativo, el cual se puede interpretar como un coeficiente de confiabilidad para el caso de que se tomen decisiones con base en una comparación del puntaje del examen global entre todos los estudiantes, tuvo un valor de 0.63. El índice absoluto, el cual es para decisiones con base en una comparación del mismo puntaje con algún criterio o estándar de desempeño, tuvo un valor de 0.55.

La figura 3 muestra gráficamente como estos índices cambiarían si se decidiese aumentar o disminuir el número de estaciones y/o el número de evaluadores por estación. Se observa que a mayor número de evaluadores y de estaciones, mayores serán los índices de generalizabilidad.

Análisis posterior relacionado con la contribución de los profesores a la varianza de las calificaciones

Ahora bien, en el estudio G llamó la atención el porcentaje de varianza relativamente alto atribuible a la faceta de los evaluadores (13%). Para entender mejor este resultado, se realizó un análisis posterior de los resultados de cada estación, el cual implica los siguientes 2 pasos: a) se dividió a los estudiantes en 3 grupos de igual tamaño, con base en su desempeño promedio en las otras 17 estaciones, y así se obtuvieron grupos de estudiantes de alto, medio y bajo rendimiento, y b) se calculó, para cada profesor participante en la estación bajo consideración, la calificación promedio en los 3 grupos de estudiantes. A pesar de que esperábamos que las diferencias entre profesores al evaluar estudiantes de un nivel de desempeño similar fuesen mínimas, en algunas estaciones resultaron amplias. Por ejemplo, en la estación 3 (fig. 4) el evaluador C otorga, en promedio, una calificación de 76% a los estudiantes de alto rendimiento que fueron examinados por él, mientras que el evaluador B es más estricto, dando una calificación de 45% a estudiantes

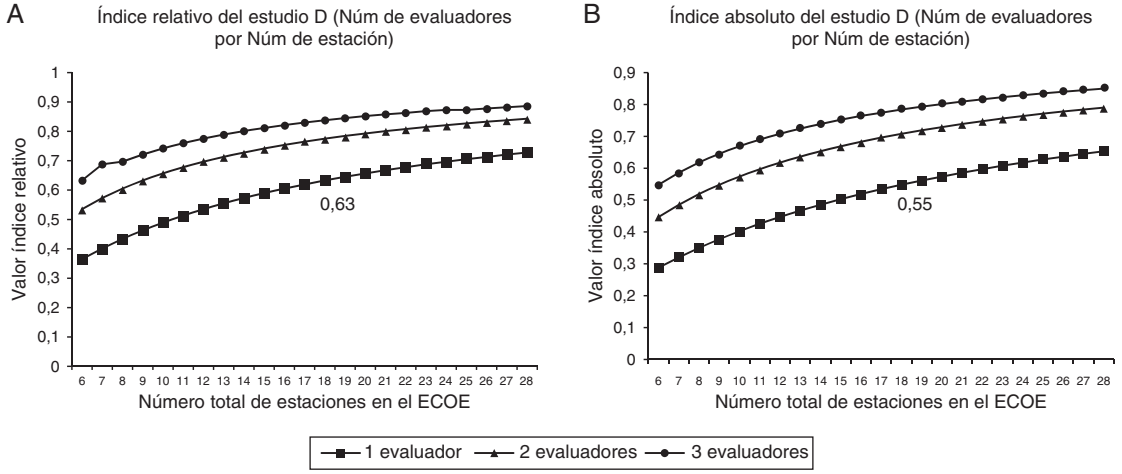

Figura 3 Índice de generalizabilidad relativo (gráfica izquierda) y absoluto (gráfica derecha), a partir del estudio D, para un ECOE-O en función del número de evaluadores y el número total de estaciones en el examen.

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

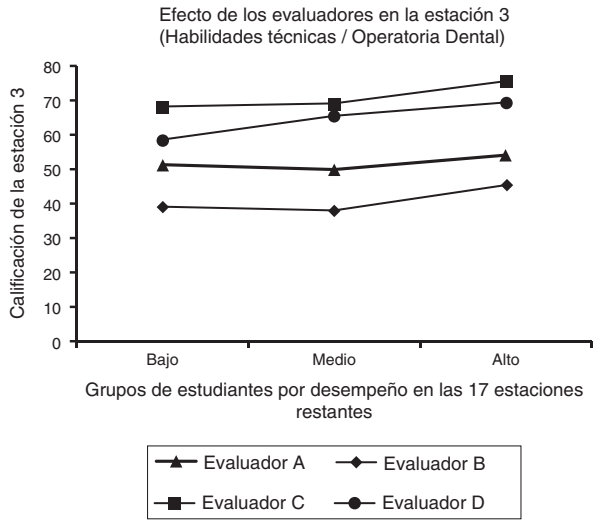

Figura 4 Efecto del evaluador por estación, en el desempeño del estudiante en el ECOE-O para la asignatura de Odontopediatría.

considerados en este mismo grupo. Incluso se observa que el evaluador C asigna una calificación de 68% a los estudiantes de bajo rendimiento, es decir, más de 20 puntos porcentuales que el evaluador B estipula para los estudiantes de alto rendimiento. Diferencias de este tipo entre los evaluadores explican el efecto relativamente grande de la faceta de los evaluadores en el estudio G.

Discusión

La confiabilidad es uno de los principales índices de calidad que aportan evidencia de validez en los resultados de una evaluación[39,40]. En este artículo presentamos los resultados de un estudio de confiabilidad, analizados a través de la teoría de la generalizabilidad, de un nuevo ECOE en odontología, los cuales se asemejan a los hallazgos obtenidos en el estudio realizado por Schoonheim-Klein et al.[30], quienes reportaron para el estudio G una varianza atribuida a los estudiantes del 6.3%. Respecto al estudio D, el índice relativo reportado fue de 0.62 y el absoluto, de 0.54. Similitudes con este estudio también se observan en el desarrollo, ya que fue administrado en más de un día y el número y duración de las estaciones fue similar.

Los índices de generalizabilidad relativo y absoluto obtenidos por Eberhard et al.[31] son relativamente más altos que los obtenidos en este estudio (0.75 y 0.69, respectivamente). Estos autores, al igual que el presente estudio, utilizaron el ECOE en una etapa preclínica; asimismo, integraron diversas áreas de la odontología como en esta investigación, que si bien se enfocó en la Odontopediatría, esta especialidad involucra diversas áreas tales como odontología restauradora, periodoncia, endodoncia y exodoncia, entre otras. A diferencia del presente estudio, llevaron a cabo diversos análisis estadísticos para obtener el índice de confiabilidad y no reportaron los resultados del estudio G de la teoría de la generalizabilidad. En nuestro estudio la fuente que contribuye principalmente a la varianza en los puntajes de los examinados es el error residual; este resultado influye en el estudio D; por tanto, puede estar relacionado directamente con la obtención de los índices relativos y absolutos de dicho estudio.

En relación con lo reportado por Näpänkangas et al.[27], es importante destacar que el sistema de evaluación clínica utilizado en dicho estudio es similar al utilizado en el ECOE del presente estudio: a través de rúbricas que evalúan el desempeño gradual de los estudiantes; los índices reportados en nuestro estudio resultan más bajos que los registrados por ellos, lo que puede relacionarse con que en Estados Unidos y Europa las competencias profesionales para los dentistas están establecidas de manera consensuada. La mayor parte de las escuelas se rigen bajo estos criterios y han trabajado de manera colaborativa para establecerlas y operacionalizarlas. En México, los planes de estudio de las escuelas de odontología son variados, y aun cuando nuestra entidad es un referente para el resto, aún falta por trabajar, de manera colegiada, lo que se busca lograr en un cirujano dentista de práctica general. Si bien se trabajó colaborativamente en la construcción del examen y se siguió rigurosamente el método, las estaciones pueden ser aún mejor definidas en cuanto a su estructuración y las competencias que evalúan. Asimismo, en este reporte no se buscó una correlación entre las evaluaciones a lo largo de todo el curso y el resultado obtenido en el ECOE de tipo sumativo. Este aspecto será reportado en nuestra investigación en un futuro.

Al comparar los resultados de la presente investigación con los de Bergus y Kreiter[32], respecto al estudio G, se observa que nuestro estudio reporta un porcentaje ligeramente menor (9.7% contra 6%) en relación con la influencia de los estudiantes en la calificación obtenida, es decir, la propia competencia clínica con la que cuentan; en cuanto a los índices obtenidos para el estudio D, reportaron un índice relativo de 0.63, semejante al reportado en este trabajo. Cabe destacar que ambos estudios tuvieron un enfoque de evaluación formativa, pero la estructura del examen fue distinta, ya que el estudio de Bergus y Kreiter contempla diversos casos de las 5 principales prácticas clínicas que desarrollan a lo largo del tercer año de medicina, que fueron acumulativos para, finalmente, darles un enfoque de evaluación sumativa. En nuestro estudio, los casos (estaciones) son representativos del curso completo de Odontopediatría.

Respecto al estudio reportado por Trejo et al.[33] en México analizado con la teoría de la generalizabilidad, se observan índices muy elevados (0.81 a 0.93, en 7 generaciones en las que se aplicó el ECOE en medicina) en relación con los reportados en la presente investigación y en otras investigaciones en las que se utilizó la teoría de la generalizabilidad para el análisis de la confiabilidad[30-32].

Pueden existir varias razones por las que los índices en el estudio actual resultaron más bajos que en algunos estudios consultados: primero, el grupo de estudiantes es más homogéneo respecto de su competencia clínica. Los estudiantes cursaban el penúltimo año de la licenciatura al momento del estudio; ellos han compartido varios años de estudio, por lo cual se puede esperar que su competencia clínica efectivamente sea más homogénea. Además, la muestra de estudiantes se sacó de 4 (de 15) grupos que constituyen la generación completa, lo cual conlleva un efecto homogeneizador en su competencia clínica. Segundo, las estaciones difieren mucho en grado de dificultad (fig. 1). Esto problematiza la generalizabilidad absoluta (expresado por el

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

índice de generalizabilidad absoluto del estudio D). Tercero, el estudio G mostró que los evaluadores difieren mucho en el grado de severidad con el que otorgan los puntajes en las estaciones. Al respecto, Park et al.[41] estudiaron la influencia del tipo de examinador en los puntajes de los estudiantes de un ECOE en odontología y encontraron que los profesores de medio tiempo tienden a evaluar con puntajes más elevados que aquellos quienes son de tiempo completo o residentes de posgrado. Esto puede estar relacionado con el grado de exigencia que estos 2 últimos solicitan y, en ocasiones, con el perder de vista que evalúan a estudiantes de pregrado en proceso de formación. En nuestro estudio participaron en su mayoría profesores especialistas en Odontopediatría, pero también dentistas de práctica general; algunos con apenas 5 años de experiencia docente y otros con más de 25 años como profesores; esto refleja una heterogeneidad entre evaluadores.

Al analizar los índices de generalizabilidad reportados en la literatura, incluidos los de este estudio, destaca que, en un porcentaje considerable, no son mayores a 0.8. En revisiones sistemáticas de los índices de confiabilidad obtenidos para este tipo de exámenes[42,43] se ha señalado que la competencia clínica es un constructo complejo, lo que dificulta obtener un índice de confiabilidad elevado, ya que son diversos los factores que influyen en su operatividad y en las calificaciones de los estudiantes. Asimismo, se ha descrito que los índices en el rango de 0.6 a 0.8 en ECOE se consideran aceptables. De manera particular, en pruebas basadas en prácticas en muestras pequeñas como la de este estudio, y en las que se utilizan examinadores, es probable obtener índices cercanos al límite inferior del rango mencionado[25,44]. Esto difiere con la literatura que reporta que el valor de 0.8 es visto como el mínimo requerido para una medición confiable[2,19].

Este reporte es parte de una investigación más amplia que abarca diferentes análisis de los resultados a partir del modelo de validez que contempla múltiples recursos de evidencia de las 5 grandes fuentes de validez señaladas en la introducción[45].

Hasta donde pudimos investigar, es la primera ocasión que en México se lleva a cabo una evaluación del tipo ECOE en el área odontológica. Por los resultados empíricos obtenidos, se han contemplado las siguientes modificaciones al ECOE propuesto en esta investigación:

- Acumular suficiente evidencia de validez en el diseño de las estaciones al revisarlas nuevamente y ajustarlas cuando sea necesario; revisar detenidamente los ítems que constituyen cada una; valorar su estructura y la relevancia de incluirlos en pruebas posteriores.
- Identificar a los profesores que han seguido adecuadamente los niveles de los criterios establecidos en las rúbricas y colocarlos como evaluadores en aquellas estaciones de mayor complejidad, como las dinámicas (de procedimiento y/o con pacientes estandarizados); mostrar con anticipación a los profesores la estación en la que estarán como evaluadores con el objeto de que se familiaricen y que, en caso de que exista alguna duda, puedan explicitarla a tiempo; y calibrar a los profesores utilizando la rúbrica con antelación en sus horarios de clase.

Una de las limitaciones de este estudio es que en el análisis, al obtener los estadísticos, los efectos de las diversas facetas están confundidos: No fue posible separar el efecto de la estación del efecto de la interacción entre estudiante y evaluador (debida a, por ejemplo, que el profesor y el estudiante se conocen de cursos pasados).

La interacción entre estación y evaluador también se puede confundir, ya que probablemente existen evaluadores más hábiles o expertos en la evaluación de ciertos dominios de la competencia clínica y, por el contrario, otros que quizá evaluaron alguna estación en la que no cuentan con la habilidad de evaluar determinados criterios. Otra de las limitaciones de este estudio es el tamaño pequeño de la muestra utilizado, lo cual conlleva que las estimaciones son relativamente imprecisas.

El diseño del estudio es trascendental para un análisis con la teoría de la generalizabilidad. Lo ideal es un estudio cruzado, en el que todos los evaluadores participan en todas las estaciones evaluando a todos los estudiantes, situación que desde los puntos de vista logístico y de factibilidad resulta imposible; sin embargo, para estudios posteriores sería interesante considerar diseños alternativos que permitan separar de mejor forma las distintas fuentes de variación en los datos del ECOE.

Conclusiones

El ECOE debe diseñarse con rigor metodológico para la interpretación y el uso adecuado de sus resultados. Debe acumularse suficiente evidencia de validez a través de diversos recursos, especialmente los relacionados con la estructura interna del instrumento, en la que se contempla la confiabilidad, analizada en este estudio.

Al momento del diseño del ECOE deben considerarse la elaboración de las estaciones, la capacitación de los evaluadores y aspectos que puedan interferir con la óptima aplicación de este sistema de evaluación, para minimizar al máximo las fuentes de variación irrelevantes que interfieran con el desempeño de los estudiantes.

La experiencia de implementar un sistema de evaluación como el ECOE en el ámbito odontológico por primera vez en México abre la puerta para continuar con pruebas y ajustes relacionados con este método. Aunque hay una diversidad de reportes en cuanto a índices de confiabilidad por debajo de 0.8 para este instrumento de evaluación, se considera el estándar de oro de la evaluación de la competencia clínica y su diseño debe ajustarse al contexto y a las condiciones del lugar donde se aplique. Su uso en nuestra entidad académica responde a una necesidad urgente de implementar un sistema de evaluación integral que evalúe objetivamente la competencia clínica de nuestros estudiantes.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Financiamiento

CONACyT, Facultad de Odontología y Facultad de Medicina, UNAM.

Autoría

OEV concibió y diseñó el proyecto, escribió la primera versión del manuscrito.
AMG concibió y diseñó el proyecto, colaboró en la escritura y revisión del manuscrito.
MSM contribuyó en el diseño del proyecto, colaboró en la escritura y revisión del manuscrito.
IL diseñó y realizó el análisis estadístico para la obtención de los datos, contribuyó en la escritura y revisión del manuscrito.
Todos los autores aportaron fuentes de literatura, realizaron la revisión crítica del artículo y aprobaron su versión final.

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

Agradecimientos

A los profesores de las asignaturas de Odontopediatría y Clínica Integral de Niños y Adolescentes de la Facultad de Odontología que participaron en este proyecto, a los pacientes estandarizados, así como a las autoridades y personal de la Secretaría de Educación Médica de la Facultad de Medicina y de la Facultad de Odontología de la UNAM.

Referencias

1. Davenport ES, Davis JE, Cushing AM, Holsgrove GJ. An innovation in the assessment of future dentists. Br Dent J. 1998;184:192–5.
2. Waas V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357:945–9.
3. Epstein RM. Assessment in medical education. N Engl J Med. 2007;356:387–96.
4. Martínez GA, Trejo MJA, Fortoul GT, Flores HF, Morales LS, Sánchez MM. Evaluación diagnóstica de conocimientos y competencias en estudiantes de medicina al término del segundo año de la carrera: el reto de construir el avión mientras vuela. Gac Med Mex. 2014;150:35–48.
5. Arnold, Walmsley RC. The use of the OSCE in postgraduate education. Eur J Dent Edu. 2008;12:16–30.
6. Schoonheim-Klein M, Walmsley AD, Habets L, van der Velden U, Manogue M. An implementation strategy for introducing an OSCE into a dental school. Eur J Dent Educ. 2005;9:143–9.
7. Durante M, Lozano S, Martínez G, Morales L, Sánchez M. Evaluación de competencias en Ciencias de la Salud. Médica Panamericana. 2010:20–6.
8. Kramer GA, Albino JEN, Andrieu SC, Hendricson WD, Henson L, Horn BD, et al. Dental student assessment toolbox. J Dent Educ. 2009;73:12–35.
9. Harden R, Stevenson M, Downie WW, Wilson GM. Clinical competence in using objective structured examination. Br Med J. 1975;1:447–51.
10. Harden RM, Gleeson FA. ASME medical educational booklet no 8: Assessment of medical competence using an objective structured clinical examination (OSCE). J Med Educ. 1979;13:41–54.
11. Näpänkangas R, Harila V, Lahti S. Experiences in adding multiple-choice questions to an objective structural clinical examination (OSCE) in undergraduate dental education. Eur J Dent Educ. 2012;16:146–50.
12. Trejo A, Blee G, Peña J. Elaboración de estaciones para el examen clínico objetivo estructurado (ECOE). Inv Ed Med. 2014;3:56–9.
13. Bhowate R, Panchbhai A, Vagha S, Tankhiwale S. Introduction of objective structured clinical examination (OSCE) in dental education in India in the subject of oral medicine and radiology. J Educ Ethics Dent. 2014;4:23–7.
14. Trejo MJA, Martínez GA, Méndez RI, Morales LS, Ruiz PL, Sánchez MM. Evaluación de la competencia clínica con el examen clínico objetivo estructurado en el internado médico de la Universidad Nacional Autónoma de México. Gac Med Mex. 2014;150:8–17.
15. Manogue M, Brown G. Developing and implementing an OSCE in dentistry. Eur J Dent Educ. 1998;2:51–7.
16. Mossey PA, Newton JP, Stirrups DR. Scope of the OSCE in the assessment of clinical skills in dentistry. Br Dent J. 2001;190:323–6.
17. Mendel N, Fuks J, Levy T, Fernández M, de Preliasco VF, Amantea A. Examen clínico objetivo y estructurado (ECOE): una propuesta innovadora en la evaluación de la Odontopediatría. Revista de la Facultad de Odontología (UBA). 2005;30:31–6.
18. Hodder RV, Rivington RN, Calcutt LE, Hart IR. The effectiveness of immediate feedback during the Objective Structured Clinical Examination. Med Educ. 1989;23:184–8.
19. Harden R, Lilley P, Patrício M. The Definitive Guide to the OSCE. The Objective Structured Clinical Examination as a Performance Assessment. Elsevier; 2016. p. 149–58, 176-179.
20. Downing SM. Validity: On the meaningful interpretation of assessment data. Med Educ. 2003;37:830–7.
21. Sánchez MM. La calidad del proceso de evaluación para la certificación del médico especialista. México: Comité Normativo Nacional de Consejos de Especialidades Médicas; 2007 [consultado 28 Abr 2015]. Disponible en: http://www.conacem.org.mx/assets/boletin_calidad.pdf
22. Wetzel AP. Factor analysis methods and validity evidence: A review of instrument development across the medical education continuum. Acad Med. 2011;87:1060–9.
23. Delgado AR, Prieto G. Fiabilidad y validez. Papeles del Psicólogo 20103167-74 [consultado 25 Mar 2016]. Disponible en: http://www.redalyc.org/articulo.oa?id=77812441007
24. Webb NM, Shavelson RJ, Haertel EH. Reliability Coefficients and Generalizability Theory. En: Rao CR, Sinharay S, editores. Handbook of Statistics 26. USA: Elsevier; 2007. p. 81–120.
25. Brown G, Manogue M, Martin M. The validity and reliability of an OSCE in dentistry. Eur J Dent Educ. 1999;3:117–25.
26. Gerrow JD, Murphy HJ, Boyd MA, Scott DA. Concurrent validity of written and OSCE components of the Canadian dental certification examinations. J Dent Educ. 2003;67:896–901.
27. Näpänkangas R, Karaharju-Suvanto T, Pyörälä E, Harila V, Ollila P, Lähdesmäki R, et al. Can the results of the OSCE predict the results of clinical assessment in dental education? Eur J Dent Educ. 2016;20:3–8.
28. Schuwirth L, van der Vleuten C. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teacher. 2011;33:783–97.
29. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: A key to unlock professional assessment. Med Edu. 2002;36:9972–8.
30. Schoonheim-Klein M, Mujitens A, Habets L, Manogue M, van der Vleuten C, Hoogstraten J, et al. On the reliability of a

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001

dental OSCE, using SEM: Effect of different days. Eur J Dent Educ. 2008;12:131–7.
31. Eberhard L, Hassel A, Bäumer A, Becker F, Beck-Mubotter J, Bömicke W, et al. Analysis of quality and feasiability of an objective structured clinical examination (OSCE) in preclinical dental education. Eur J Dent Educ. 2011;15:172–8.
32. Bergus GR, Kreiter CD. The reliability of summative judgements based on objective structured clinical examination cases distributed across the clinical year. Med Edu. 2007;41:661–6.
33. Trejo MA, Sánchez MM, Méndez RI, Martínez GA. Reliability analysis of the objective structured clinical examination using generalizability theory. Med Educ Online. 2016;21:31650, http://dx.doi.org/10.3402/meo.v21.31650.
34. Taguchi N, Ogawa T. OSCEs in Japanese postgraduate clinical training Hiroshima experience 2000-2009. Eur J Dent Educ. 2010;14:203–9.
35. Larsen T, Jeppe-Jensen D. The introduction and perception of an OSCE with an element of self-and peer-assessment. Eur J Dent Educ. 2008;12:2–7.
36. Schoonheim-Klein M, Habets L, Artman A, van der Vleuten C, Hoogstraten J, van der Velden U. Implementing an Objective Structured Clinical Examination (OSCE) in dental education: Effects on students' learning strategies. Eur J Dent Educ. 2006;10:226–35.
37. Schoonheim-Klein M, Mujitens A, Habets L, Manogue M, van der Vleuten C. Who will pass the dental OSCE? Comparison of the Angoff and the borderline regression standard setting methods. Eur J Dent Educ. 2009;13:162–71.
38. Brennan RL. Generalizability Theory. En: Instructional Topics in Educational Measurement. Iowa: National Council of Medical Education; 1992. p. 27–34.
39. Downing SM. Reliability: On the reproducibility of assessment data. Med Educ. 2004;38:1006–12.
40. Downing SM, Yudkowsky R. Reliability. En: Downing SM, Yudkowsky R, editores. Assessment in Health Professions Education. New York, London: Routdlege. Taylor and Francis Group; 2009. p. 57–74.
41. Park SE, Kim A, Kristiansen J, Karimbux NY. The influence of examiner type on dental students' OSCE scores. J Dent Edu. 2015;79:89–94.
42. Patrício M, Juliao M, Fareleira F, Vaz A. Is the OSCE a feasible tool to asses competencies in undergraduate medical education. Med Teacher. 2013;35:503–14.
43. Brannick MT, Tugba EH, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Med Educ. 2011;45:1181–9.
44. Youngman MB. Analysing social and research education data. En: Brown G, Manogue M, Martin M, editors. The validity and reliability of an OSCE in dentistry, 3. Eur J Dent Educ; 1999. p. 117–25.
45. Espinosa V.O., Martínez G.A., Sánchez M.M. Evidencias de validez para el Examen Clínico Objetivo Estructurado. Cartel y presentación oral. XX Conferencia Panamericana de Educación Médica. V Congreso Internacional de Educación Médica. IV Congreso Internacional de Simulación. Cancún, México, 2016.

Cómo citar este artículo: Espinosa-Vázquez O, et al. Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. Inv Ed Med. 2016. http://dx.doi.org/10.1016/j.riem.2016.09.001